

**PATENT APPLICATION**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of

Taro TERAOKA

Application No.: 10/658,810

Filed: September 10, 2003

Docket No.: 117127

For: INFORMATION PROCESSING SYSTEM

**CLAIM FOR PRIORITY**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested for the above-identified patent application and the priority provided in 35 U.S.C. §119 is hereby claimed:

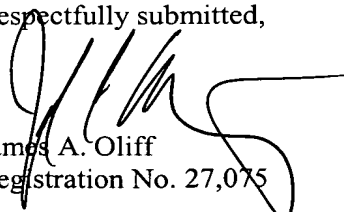
Japanese Patent Application No. 2003-037440 filed February 14, 2003

In support of this claim, a certified copy of said original foreign application:

☒ is filed herewith.

It is requested that the file of this application be marked to indicate that the requirements of 35 U.S.C. §119 have been fulfilled and that the Patent and Trademark Office kindly acknowledge receipt of this document.

Respectfully submitted,

  
James A. Oliff  
Registration No. 27,075

Joel S. Armstrong  
Registration No. 36,430

JAO:JSA/mlb

Date: December 30, 2003

**OLIFF & BERRIDGE, PLC**  
**P.O. Box 19928**  
**Alexandria, Virginia 22320**  
**Telephone: (703) 836-6400**

DEPOSIT ACCOUNT USE  
AUTHORIZATION  
Please grant any extension  
necessary for entry;  
Charge any fee due to our  
Deposit Account No. 15-0461

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 2月14日  
Date of Application:

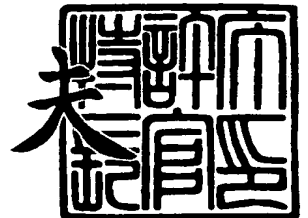
出願番号 特願2003-037440  
Application Number:  
[ST. 10/C]: [JP 2003-037440]

出願人 富士ゼロックス株式会社  
Applicant(s):

2003年11月11日

特許庁長官  
Commissioner,  
Japan Patent Office

今井 康夫



出証番号 出証特2003-3093042

【書類名】 特許願

【整理番号】 FE02-02232

【提出日】 平成15年 2月14日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/00

【発明者】

【住所又は居所】 神奈川県足柄上郡中井町境 4 3 0 グリーンテクなかい  
富士ゼロックス株式会社内

【氏名】 寺尾 太郎

【特許出願人】

【識別番号】 000005496

【氏名又は名称】 富士ゼロックス株式会社

【代理人】

【識別番号】 110000154

【氏名又は名称】 特許業務法人はるか国際特許事務所

【代表者】 金山 敏彦

【電話番号】 03-5367-2791

【手数料の表示】

【予納台帳番号】 185835

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報処理システム

【特許請求の範囲】

【請求項 1】 処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持し、処理対象データの要求として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供するサーバコンピュータシステムと、

前記サーバコンピュータシステムに対して通信可能に接続され、処理対象となるデータを前記サーバコンピュータシステムから取得するクライアント側システムと、

を含んでなることを特徴とする情報処理システム。

【請求項 2】 処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持する保持手段と、

処理対象データを要求する情報として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供する提供手段と、

を含むことを特徴とするコンピュータシステム。

【請求項 3】 請求項 2 に記載のコンピュータシステムにおいて、  
処理対象となるデータに基づいて特徴量を演算する演算手段をさらに含み、  
前記演算手段は、  
処理対象となるデータを所定サイズのデータフラグメントの列に分割し、  
データフラグメントごとに、各データフラグメントに基づく特徴量を演算し、  
前記保持手段にデータフラグメントと特徴量とを関連づけて保持させ、  
データフラグメントの列に対応する特徴量の列を生成し、  
当該特徴量の列に基づく特徴量を演算し、前記保持手段に当該特徴量の列と、  
それに基づいて演算された特徴量とを関連づけて保持させる  
ことを特徴とするコンピュータシステム。

【請求項 4】 請求項 3 に記載のコンピュータシステムにおいて、  
前記演算手段は、特徴量の列に基づく特徴量を演算する際に、特徴量の列に含

まれる各特徴量ごとに繰り返し演算により、当該特徴量の列に基づく特徴量を演算しており、

特徴量の列がN個の特徴量からなる場合は、1からN-1個までの特徴量に対する前記繰り返し演算の結果を前記保持手段に保持させることを特徴とするコンピュータシステム。

【請求項5】 請求項2に記載のコンピュータシステムにおいて、  
処理対象となるデータに基づいて特徴量を演算する演算手段をさらに含み、  
前記演算手段は、  
処理対象となるデータを所定サイズのデータフラグメントの列に分割し、  
データフラグメントごとに、各データフラグメントに基づく特徴量を演算し、  
前記所定サイズと、演算された特徴量のサイズとを比較し、前記所定サイズが特徴量のサイズより小さい場合は、データフラグメントそのものを前記保持手段に保持させ、前記所定サイズが、前記特徴量のサイズより大きい場合は、前記保持手段にデータフラグメントと特徴量とを関連づけて保持させ、  
データフラグメントの列に対応する、特徴量を含む列を生成し、  
当該特徴量を含む列に基づく特徴量を演算し、前記保持手段に当該特徴量を含む列と、それに基づいて演算された特徴量とを関連づけて保持させる  
ことを特徴とするコンピュータシステム。

【請求項6】 請求項2に記載のコンピュータシステムにおいて、  
前記保持手段は、少なくとも一つの特徴量を含んでなる特徴量セットに基づいて演算された特徴量を保持し、  
前記提供手段は、処理対象データを要求する情報として受け入れた特徴量が、特徴量セットに関連づけられているときには、当該特徴量セットに含まれる各特徴量を要求元に提供する

ことを特徴とするコンピュータシステム。

【請求項7】 請求項2から6のいずれか一項に記載のコンピュータシステムにおいて、

前記保持手段は、前記特徴量に関連づけて、当該特徴量の演算方法を特定する特徴量演算方法特定情報を保持することを特徴とするコンピュータシステム。

【請求項 8】 請求項 2 から 7 のいずれか一項に記載のコンピュータシステムにおいて、

前記保持手段は、前記特徴量に、当該特徴量の演算時点における所定の演算の状況に関連する情報を含めて保持していることを特徴とするコンピュータシステム。

【請求項 9】 コンピュータシステムを用い、

処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持する工程と、

処理対象データを要求する情報として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供する工程と、

を実行することを特徴とする情報処理方法。

【請求項 10】 コンピュータシステムに、

処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持する手順と、

処理対象データを要求する情報として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供する手順と、

を実行させることを特徴とするプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、処理対象となる情報を保持し、提供する情報処理システムに関する。

【0002】

【従来の技術】

コンピュータ技術の発展、及びその社会への普及の度合いは近年ますます進行し、もはやコンピュータ制御されない装置は存在しないといっても過言でないほどである。こうした状況下、各コンピュータを利用する利用者、すなわちエンド

ユーザ、プログラム開発者、情報提供者等は、コンピュータ内で処理の対象となるデータを特定するため、通常、ファイル名等の情報を用いている。

#### 【0003】

コンピュータ内部では、ファイル名は、対応するポインタ（例えば、インデックスノード、すなわち i ノード）に変換され、ディスク上の当該ポインタで示される位置から、当該ファイル名に対応するデータが読み出される。こうした従来一般的な方法については、例えば非特許文献 1 が詳しい。

#### 【0004】

##### 【非特許文献 1】

Maurice J. Bach 著，坂本文・多田好克・村井純訳，UNIX カーネルの設計，1991 年 6 月 10 日，共立出版株式会社 初版発行

#### 【0005】

##### 【発明が解決しようとする課題】

しかしながら、上記従来のような処理対象データの提供と取得の方法では、ユーザは、ファイル名を考え出さねばならず、ユーザにとって利便性が低い。またデータの先頭部分を基準にファイル名を生成する装置も存在するが、例えば報告書を多く作成するユーザの場合、「report」などで始まる名称が多くなるなど、利便性が低い。

#### 【0006】

本発明は上記実情に鑑みて為されたもので、ユーザにとっての利便性が高い情報処理システムを提供することをその目的の一つとする。

#### 【0007】

##### 【課題を解決するための手段】

上記従来例の問題点を解決するための本発明は、情報処理システムにおいて、処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持し、処理対象データの要求として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供するサーバコンピュータシステムと、前記サーバコンピュータシステムに対して通信可能に接続され、処理対象となるデータを前記サーバコンピュータシステムから

取得するクライアント側システムと、を含んでなることを特徴としている。

【0008】

また、上記従来例の問題点を解決するための本発明は、コンピュータシステムであって、処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持する保持手段と、処理対象データを要求する情報として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供する提供手段と、を含むことを特徴としている。

【0009】

ここでさらに処理対象となるデータに基づいて特徴量を演算する演算手段をさらに含み、前記演算手段は、処理対象となるデータを所定サイズのデータフラグメントの列に分割し、データフラグメントごとに、各データフラグメントに基づく特徴量を演算し、前記保持手段にデータフラグメントと特徴量とを関連づけて保持させ、データフラグメントの列に対応する特徴量の列を生成し、当該特徴量の列に基づく特徴量を演算し、前記保持手段に当該特徴量の列と、それに基づいて演算された特徴量とを関連づけて保持させることとするのも好ましい。

【0010】

さらに、前記演算手段は、特徴量の列に基づく特徴量を演算する際に、特徴量の列に含まれる各特徴量ごとに繰り返し演算により、当該特徴量の列に基づく特徴量を演算しており、特徴量の列がN個の特徴量からなる場合は、1からN-1個までの特徴量に対する前記繰り返し演算の結果を前記保持手段に保持させることとしてもよい。

【0011】

また、処理対象となるデータに基づいて特徴量を演算する演算手段をさらに含み、前記演算手段は、処理対象となるデータを所定サイズのデータフラグメントの列に分割し、データフラグメントごとに、各データフラグメントに基づく特徴量を演算し、前記所定サイズと、演算された特徴量のサイズとを比較し、前記所定サイズが特徴量のサイズより小さい場合は、データフラグメントそのものを前記保持手段に保持させ、前記所定サイズが、前記特徴量のサイズより大きい場合は、前記保持手段にデータフラグメントと特徴量とを関連づけて保持させ、デー



タフラグメントの列に対応する、特徴量を含む列を生成し、当該特徴量を含む列に基づく特徴量を演算し、前記保持手段に当該特徴量を含む列と、それに基づいて演算された特徴量とを関連づけて保持させることとするのも好ましい。

#### 【0012】

さらに、前記保持手段は、少なくとも一つの特徴量を含んでなる特徴量セットに基づいて演算された特徴量を保持し、前記提供手段は、処理対象データを要求する情報として受け入れた特徴量が、特徴量セットに関連づけられているときには、当該特徴量セットに含まれる各特徴量を要求元に提供することとするのも好ましい。

#### 【0013】

さらに、前記保持手段は、前記特徴量に関連づけて、当該特徴量の演算方法を特定する特徴量演算方法特定情報を保持することとしてもよいし、前記保持手段は、前記特徴量に、当該特徴量の演算時点における所定の演算の状況に関連する情報を含めて保持していることとしてもよい。

#### 【0014】

また、本発明のある態様に係る方法は、コンピュータシステムを用い、処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持する工程と、処理対象データを要求する情報として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供する工程と、を実行することを特徴としている。

#### 【0015】

また、本発明の別の態様に係るプログラムは、コンピュータシステムに、処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持する手順と、処理対象データを要求する情報として特徴量を受け入れ、前記保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供する手順と、を実行させることを特徴としている。

#### 【0016】

なお、このプログラムは、コンピュータ可読な記録媒体に格納されてもよい。

#### 【0017】

**【発明の実施の形態】**

本発明の実施の形態について図面を参照しながら説明する。本発明の実施の形態に係る情報処理システムは、図1に示すように、サーバコンピュータシステム1と、クライアント側コンピュータシステム2とを含んで構成される。サーバコンピュータシステム1は、データストレージ部11と、制御部12と、記憶部13と、通信制御部14とを主として含み、クライアント側コンピュータシステム2は、パーソナルコンピュータ等を含んで構成される。サーバコンピュータシステム1とクライアント側コンピュータシステム2とは、それぞれ少なくとも一つのコンピュータを含んで構成されるコンピュータシステムである。

**【0018】**

これらサーバコンピュータシステム1と、クライアント側コンピュータシステム2とは、互いにネットワークを介して接続されている。ここでネットワークは、シリアル伝送回線（USBやIEEE1394等を含む）、電気通信回線（イーサネット（登録商標）その他）等、種々のデータ通信経路のことである。

**【0019】**

サーバコンピュータシステム1のデータストレージ部11は、クライアント側コンピュータシステム2の処理対象となるデータの実体と、そのデータの実体に基づいて演算された特徴量とを保持している。具体的にデータストレージ部11は、基本的には、図2（a）に示すように、データの実体とそのデータ実体に基づいて演算される特徴量とを関係づけて記憶している。

**【0020】**

制御部12は、記憶部13に格納されているプログラムに従って動作している。この制御部12は、通信制御部14を介してクライアント側コンピュータシステム2から受信される指示に基づき、処理対象となるデータ実体に対する特徴量を演算する処理（演算処理）を実行し、そのデータ実体と特徴量とを関連づけてデータストレージ部11に格納する。また、この制御部12は、クライアント側コンピュータシステム2から特徴量とともに当該特徴量に対応するデータの要求を受信すると、当該特徴量に関連づけられているデータ実体をデータストレージ部11から検索し、検索の結果、当該特徴量に関連づけられているデータ実体が

データストレージ部 11 から見いだされたときには、当該データ実体をクライアント側コンピュータシステム 2 に対して送信するよう、通信制御部 14 に指示する（提供処理）。

#### 【0021】

これら制御部 12 が行う演算処理と提供処理とのより詳しい説明、並びにその変形例については後に詳しく述べる。

#### 【0022】

記憶部 13 は、制御部 12 が実行するプログラムを保持するディスク装置を含み、また、制御部 12 が、その処理の過程で生成したデータを記憶するためのワークメモリとしての RAM（Random Access Memory）等を含む。通信制御部 14 は、ネットワークに接続されており、ネットワークを介して到来するデータや要求等を受信して制御部 12 に出力する。また、この通信制御部 14 は、制御部 12 から入力される指示に従って、ネットワークを介してデータ等をクライアント側コンピュータシステム 2 等へ送信する。なお、通信制御部 14 は、受信したデータ等の送信元を特定する情報を制御部 12 に出力し、制御部 12 は、当該情報によって通信制御部 14 に対して送信先を特定する等の処理を行うが、これらは広く知られた通信方式であるので、その詳しい説明は省略する。

#### 【0023】

ここで制御部 12 の処理の具体的内容と、その変形例について説明する。まず、制御部 12 の演算処理について説明する。制御部 12 は、処理対象となったデータに基づいて、予め定められた方法によって特徴量としての値を演算し、当該データと、特徴量となった値とを関連づけてデータストレージ部 11 に格納する（図 2（a））。ここで特徴量は、例えばデータに基づくハッシュ値（適当な暗号学的ハッシュ関数により得られた値である。暗号学的ハッシュ関数としては、例えば SHA-1 等がある。）などであってよいが、任意のランダム関数によって演算されるものであればいかなるものであってもよい。すなわち、ここでの特徴量は、データの各々に固有のものとなれば問題がない。さらに言えば、特徴量は、十分広い空間上に疎に分布し、データ（コンテンツ）から自動的に定まるデータ（コンテンツ）の識別子であればよい。

以下の例では、20バイトのハッシュ値であるとして説明をするが、これに限られるものではない。

#### 【0024】

また制御部12は、演算処理として、次のような処理としてもよい。すなわち制御部12は、図3に示すように、処理対象となったデータを固定長のデータフラグメントに区切ってデータフラグメントの列とし（S1）、各データフラグメントに基づく特徴量を演算して（S2）、各データフラグメントと、それぞれに基づいて生成された特徴量とを関連づけてデータストレージ部11に格納する（S3）。さらに制御部12は、当該データフラグメント列に対応する特徴量の列に対して、特徴量（以下、これを区別のために二次特徴量と呼ぶが、特徴量と同じ方法で演算される20バイトのハッシュ値で構わない）を生成し（S4）、当該特徴量の列と、それに対する二次特徴量とを関連づけてデータストレージ部11に格納する（S5）。従って、この場合におけるデータストレージ部11には、図2（a）に示したようなもののほかに、又はそれに代わって、図2（b）に示すように、データフラグメントと、それに対する特徴量、並びに、特徴量の列に対する二次特徴量とがそれぞれ関連づけて保持されているようになる。

#### 【0025】

このようにデータをデータフラグメントに分割したときには、制御部12の提供処理と、それに対するクライアント側コンピュータシステム2の動作は、次のようになる。すなわち、クライアント側コンピュータシステム2では、所望の処理対象データに対する二次特徴量をサーバコンピュータシステム1に送信する。サーバコンピュータシステム1の制御部12は、通信制御部14を介して当該二次特徴量を受け入れて、それに対応する特徴量の列をデータストレージ部11から取得する。次に、制御部12は、当該特徴量の列の各々に対応するデータフラグメント（実体データのフラグメント）を取得し、それらを組み合わせて元のデータを生成し、通信制御部14を介して要求元であるクライアント側コンピュータシステム2に送信する。

#### 【0026】

なお、この制御部 12 の演算処理におけるデータフラグメントの生成において、処理対象のデータ末尾が、データフラグメントのサイズに満たない場合は、所定の値をパディングすればよい。

#### 【0027】

また、ここでの特徴量は、繰り返し演算により演算されてもよい。つまり、制御部 12 は例えば二次特徴量を演算する際に、図 4 に示すような処理を行う。すなわち、制御部 12 は、カウンタ  $i$  を記憶部 13 に保持して、この  $i$  を「1」にリセットし (S11)、特徴量の初期値を特徴量の現在値として記憶部 13 に格納して (S12)、特徴量の列に含まれる  $N$  個の特徴量のうち、 $i$  番目の特徴量を注目特徴量として、当該注目特徴量と、現在記憶部 13 に格納されている現在値とに基づき、現在特徴量を更新して記憶部 13 に格納する (S13)。そして制御部 12 は、 $i$  をインクリメントし (S14)、 $i$  が  $N$  を超えたか否かを判断し (S15)、超えていなければ (No ならば)、処理 S13 に戻って処理を続け、処理 S15 において  $i$  が  $N$  を超えていれば (Yes ならば)、その時点で記憶部 13 に格納されている現在値を二次特徴量として出力して (S16)、処理を終了する。

#### 【0028】

この場合、制御部 12 は、処理 S15 に先立って  $i = N$  であるか否かを判定し、 $i = N$  であれば、その時点の現在値 (1 から  $N - 1$  個までの特徴量に対する繰り返し演算の結果) を記憶部 13 に格納しておき、処理 S16 において、二次特徴量とともに当該 1 から  $N - 1$  個までの特徴量に対する繰り返し演算の結果を出力し、これらの特徴量の列に関連づけてデータストレージ部 11 に格納してもよい。

#### 【0029】

このようにすると、例えば後に、処理対象となったデータの末尾に新たなデータ部分が付加されて、処理対象となったデータが更新された場合、当該更新後のデータに対する演算処理を行う際に、前回末尾であったデータフラグメント ( $N$  個目のデータフラグメント) 以降のデータフラグメントについてのみ、それぞれの特徴量を演算するとともに、二次特徴量の演算の際には、 $N - 1$  個目までの二

次特徴量に相当する量は既に演算されて保持されているので、この相当量（前回の1からN-1個までの特徴量に対する繰り返し演算の結果）を特徴量の初期値として処理S12で設定し、処理S11で $i = N$ として、N個目のデータフラグメントに対する特徴量から演算を開始することができる。これによって演算処理の負荷が軽減される。

#### 【0030】

また、特徴量のビット数が定まっていない場合は、実体データのサイズより特徴量のサイズが大きくなったり、データフラグメントのサイズより特徴量のサイズが大きくなるなどの事態も考えられる。このような場合は、演算される特徴量の代わりに、実体データやデータフラグメントそのものを特徴量として用いることとしてもよい。データ全体のサイズが小さくなり、通信等の負荷を軽減できるからである。

#### 【0031】

この場合、制御部12は、例えばデータフラグメントごとに、各データフラグメントに基づく特徴量を演算し、データフラグメントのサイズと、演算された特徴量のサイズとを比較し、データフラグメントのサイズが特徴量のサイズより小さい場合は、データフラグメントそのものを特徴量としてデータストレージ部11に保持させる。

#### 【0032】

また、制御部12は、データフラグメントのサイズが、特徴量のサイズより大きい場合は、データストレージ部11にデータフラグメントと特徴量とを関連づけて保持させる。

#### 【0033】

さらに、二次特徴量は、一つの実体データに対する特徴量の列に対するものだけでなく、複数個の実体データに対する特徴量のセットに対するものであってもよい。さらにこのセットには別の二次特徴量が含まれてもよい。すなわち実体データAに対して、特徴量 $\alpha$ が演算され、実体データBに対して特徴量の列 $\beta$ が演算され、さらに特徴量の列 $\beta$ に対して二次特徴量 $\gamma$ が演算されているとするとき、特徴量 $\alpha$ と、二次特徴量 $\gamma$ とのセットに対してさらに、別の二次特徴量 $\epsilon$ を演

算し、この $\varepsilon$ に、 $\alpha$ と $\gamma$ とを関連づけてデータストレージ部11に格納しておいてもよい。

#### 【0034】

このようにすると、例えば $\varepsilon$ だけを知っておけば、クライアント側コンピュータシステム2から $\varepsilon$ を送信して、それに対するデータを要求すると、サーバコンピュータシステム1が $\alpha$ と $\gamma$ とを送信するので、クライアント側コンピュータシステム2では、さらに実体データAが必要な場合は、当該受信した $\alpha$ をサーバコンピュータシステム1に送信して、実体データAを要求することができるようになり、サーバコンピュータシステム1に階層的構造を有するファイルシステムを構成することができる。

#### 【0035】

なお、さらに制御部12は、各特徴量の演算方法を、各特徴量に関連づけてデータストレージ部11に格納してもよいし、また、各特徴量に、当該特徴量の演算日時や、演算を指示した人物の認証情報等、特徴量の演算時点における所定の演算の状況に関連する情報を含めることとしてもよい。このように特徴量に認証情報等を含めることにより、例えば処理対象となるデータの数に対して、特徴量がそれぞれ固有に割り当てられない場合（特徴量のビット数が少ないなど）にも、各処理対象データを相互に識別できるようになる。

#### 【0036】

本実施の形態に係る情報処理システムは、以上のように構成されており、次のように動作する。ユーザがクライアント側コンピュータシステム2で生成した処理対象データは、ネットワークを介してサーバコンピュータシステム1に送信される。

#### 【0037】

サーバコンピュータシステム1では、当該受信した処理対象データに基づく特徴量としてハッシュ値を演算し、処理対象データに関連づけて当該ハッシュ値を格納し、このハッシュ値をクライアント側コンピュータシステム2に対して送信する。

#### 【0038】

また、サーバコンピュータシステム 1 は、処理対象データが一定のサイズより長いなど、所定の条件を満足する場合には、処理対象データを所定サイズのデータフラグメントに分割して、各データフラグメントに対するハッシュ値を演算し、各データフラグメントと、それぞれに基づいて演算されたハッシュ値とを互いに関連づけて格納し、さらに、データフラグメントの順に、それぞれのハッシュ値を配列したハッシュ値の列について、二次特徴量としてのハッシュ値を演算し、ハッシュ値の列に関連づけて、当該二次特徴量としてのハッシュ値を格納し、二次特徴量をクライアント側コンピュータシステム 2 に送信する。

#### 【0039】

またクライアント側コンピュータシステム 2 は、ユーザの操作に応じて、複数のハッシュ値を指定して、それらをグループ化（一つのディレクトリにまとめる）する指示をサーバコンピュータシステム 1 に対して送信する。

#### 【0040】

すると、サーバコンピュータシステム 1 では、当該指示を受信して、その指示において指定された複数のハッシュ値をハッシュ値のセット（特徴量セット）として、それらの特徴量セットに対するハッシュ値を演算し、その特徴量セットとハッシュ値とを関連づけて記憶し、特徴量セットに対するハッシュ値をクライアント側コンピュータシステム 2 に送信する。

#### 【0041】

次に、クライアント側コンピュータシステム 2 において、実体のデータが必要となった場合は、その処理対象となった実体データに対応するハッシュ値が知られていれば、当該ハッシュ値をサーバコンピュータシステム 1 に送信すれば、サーバコンピュータシステム 1 が当該ハッシュ値に対応する実体データを要求元のクライアント側コンピュータシステム 2 に対して送信することになる。

#### 【0042】

また実体データに対応するハッシュ値が直接わからない場合も、グループ化したときの特徴量セットに対するハッシュ値が知られていれば、当該ハッシュ値をサーバコンピュータシステム 1 に送信して、サーバコンピュータシステム 1 が当該ハッシュ値に対応する特徴量セット（ハッシュ値のセット）を要求元のクライ



アント側コンピュータシステム 2 に対して送信することになるので、このセットに含まれている、所望のデータに対するハッシュ値を利用すればよい。

#### 【0043】

さらに、このような特徴量セットに対するハッシュ値（特徴量）をさらに特徴量セット（二次セット）として、この二次セットに対する特徴量をさらに演算するといったようにすると、階層的なディレクトリ構造が表現できる。このような場合、クライアント側コンピュータシステム 2 では、ルートディレクトリに相当するディレクトリ（特徴量セット）に対する特徴量だけが知られていれば済むので、ユーザは例えば 20 バイト程度の情報さえ携えていれば、サーバコンピュータシステム 1 との間で通信可能な環境下でならばどこでも、自己のデータを取り出すことができる。

#### 【0044】

また、実体データに対応するハッシュ値を送信することによって、他のユーザとの間でファイルを交換することも可能となる。なお、ここでの例のように、ハッシュ値を 20 バイト程度としておくと、当該 20 バイトで表現される値の群に対して、実際のデータは、とびとびに存在する。従って、悪意あるユーザが任意に 20 バイトの値を生成して、サーバコンピュータシステム 1 に送信したとしても、当該任意に生成された 20 バイトの値に対する実体データが存在する可能性はほとんどない。さらに、データをフラグメント化している場合は、仮に偶然対応するデータがあっても、多くの場合、そのデータフラグメントだけが取得されることとなって（データフラグメントに対応するハッシュ値の方が二次特徴量としてのハッシュ値よりも数が多いため）、意味あるデータが取得される可能性はやはり相当低いものである。

#### 【0045】

このように、本実施の形態の情報処理システムでは、情報が漏えいしてしまう可能性も極めて低い。

#### 【0046】

なお、実体データに対応するハッシュ値、又は実体データをフラグメント化したときの二次特徴量としてのハッシュ値については、元の実体データのプレビュー

ー画像データ等に関連づけておき、必要に応じてクライアント側コンピュータシステムに提供するようにしてもよい。

【0047】

以上説明したように、本実施の形態の情報処理システムでは、ユーザは処理対象のデータについて一々ファイル名を設定することなく、各処理対象データを特定し、取得することができるようになって、利便性が向上する。

【0048】

また、ここまでの説明では、サーバコンピュータシステム1とクライアント側コンピュータシステム2とが別体である例について説明したが、これらは一体のコンピュータシステムであってもよい。

【図面の簡単な説明】

【図1】 本発明の実施の形態に係る情報処理システムの構成ブロック図である。

【図2】 データストレージ部11に格納されている情報の例を表す説明図である。

【図3】 演算処理の概要を表す説明図である。

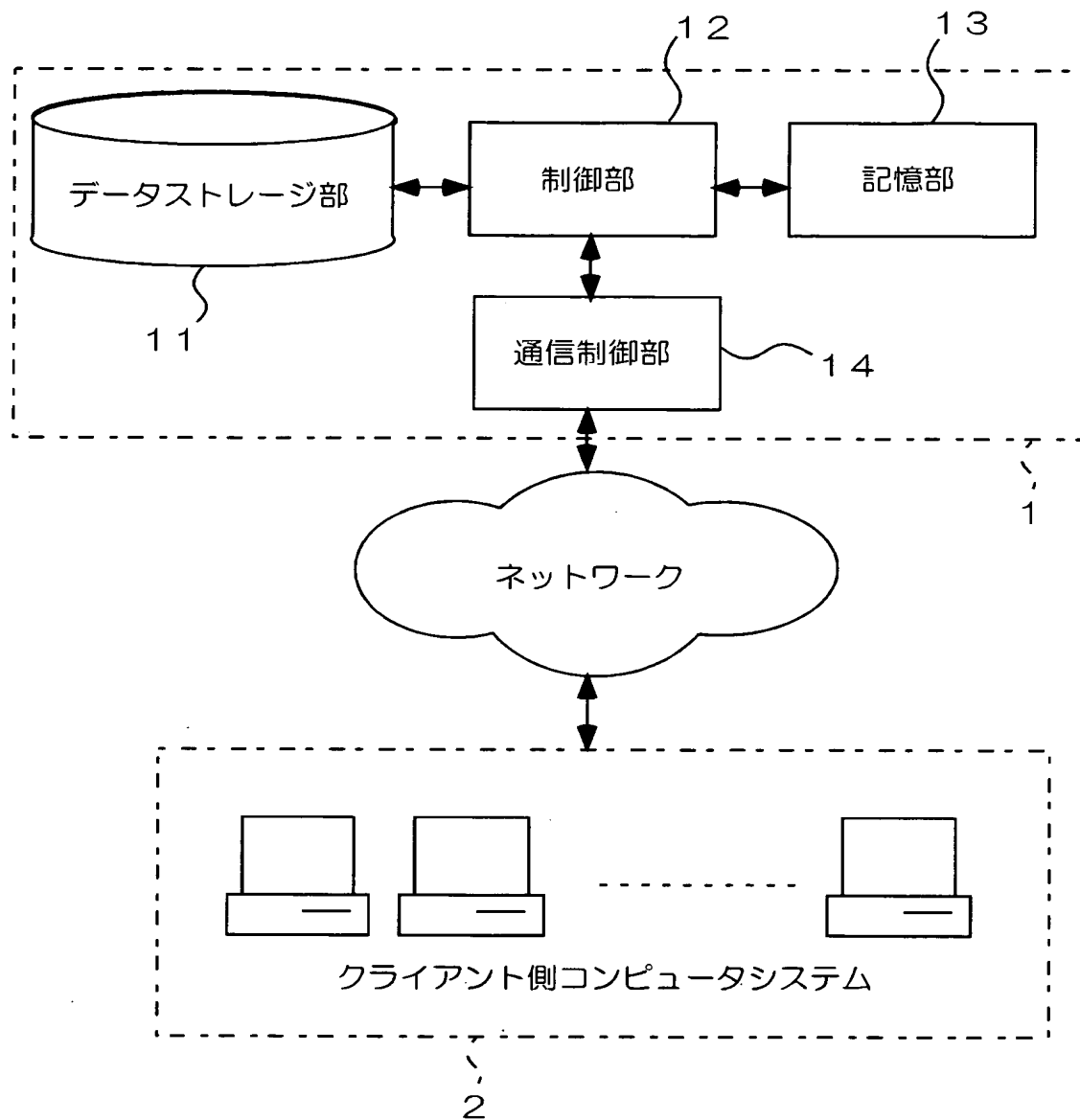
【図4】 特徴量の演算の例を表すフローチャート図である。

【符号の説明】

1 サーバコンピュータシステム、2 クライアント側コンピュータシステム、11 データストレージ部、12 制御部、13 記憶部、14 通信制御部。

【書類名】 図面

【図 1】



【図 2】

(a)

特徴量	実体データ
a a a a	b b b b b...
c c c c	d d d d d d...
⋮	⋮

(b)

特徴量	実体データ又は特徴量列
a a a a	b b b b b...
e e e e	d d d d d d...
⋮	⋮
x x x x	a a a a, e e e e...
⋮	⋮



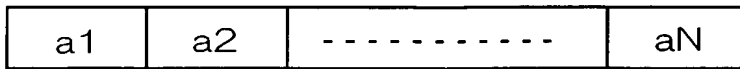
【図 3】

実体データ



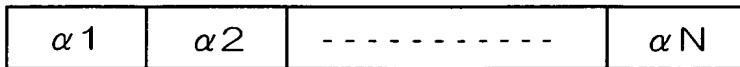
↓(S1)

データフラグメント



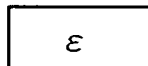
↓(S2)

特徴量



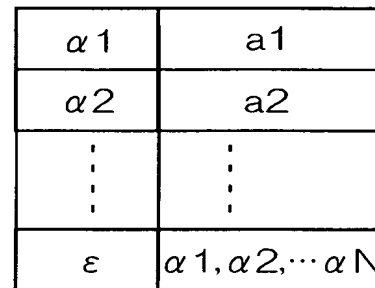
↓(S3)

二次特徴量

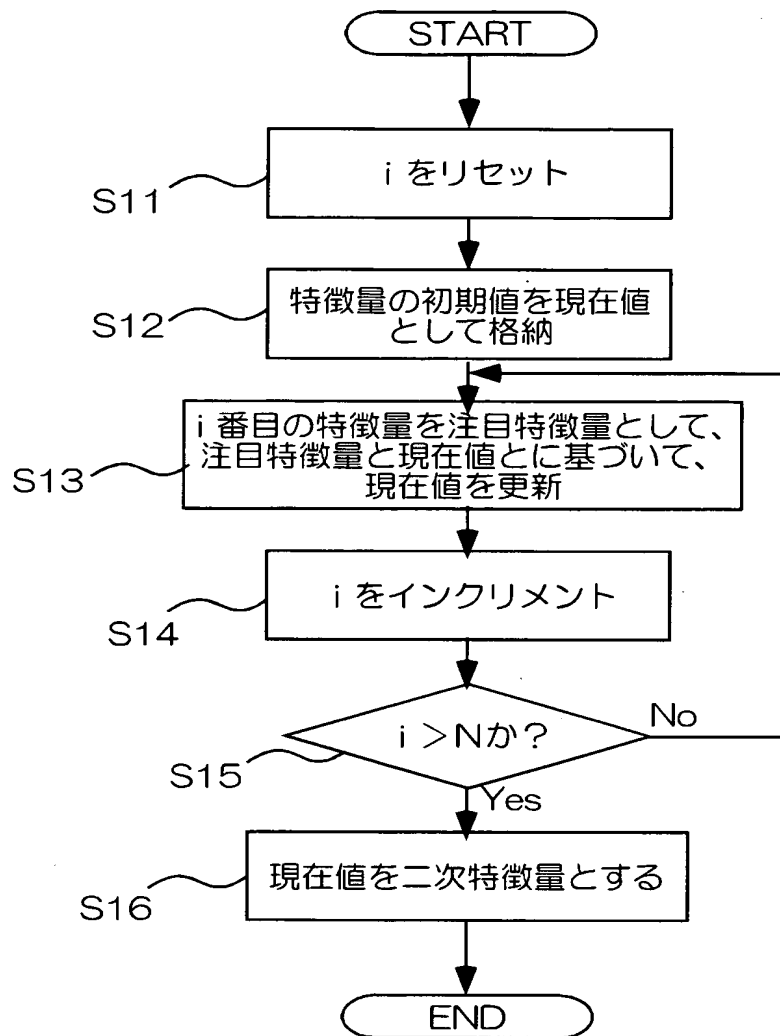


(S4)

(S5)



【図 4】



【書類名】 要約書

【要約】

【課題】 ユーザにとっての利便性が高い情報処理システムを提供する。

【解決手段】 サーバコンピュータシステム 1 が、処理対象となるデータと、当該データに基づいて演算される特徴量とを関連付けて保持し、処理対象データの要求として特徴量を受け入れ、保持しているデータのうち、受け入れた特徴量に関連付けられているデータを要求元に提供し、このサーバコンピュータシステム 1 に対して通信可能に接続され、処理対象となるデータをサーバコンピュータシステム 1 から取得するクライアント側コンピュータシステム 2 とを含む情報処理システムである。

【選択図】 図 1

特願 2 0 0 3 - 0 3 7 4 4 0

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 4 9 6 ]

1 . 変更年月日

1 9 9 6 年    5 月 2 9 日

[変更理由]

住所変更

住    所

東京都港区赤坂二丁目 1 7 番 2 2 号

氏    名

富士ゼロックス株式会社